

STAT

Page Denied

RESTRICTED

STAT

PROCEDURE FOR OBJECTIVELY EVALUATING THE
ACCURACY OF FORECASTS, ACCORDING TO A UNIFORM SCALE

K. V. Brodovitskiy

To appear in the original

Author's Abstract

The introduction of stochastically founded concepts of "difficulties" and "facilities" in forecasting enables us to establish a universal scale of evaluation. Zero on the standardized scale designates the failure of the evaluation method of forecasting; null evaluations correspond to cases of stochastic independence of the forecast from predicted phenomena. The selection of the graduation of the upper and lower halves of the scale realizes the requirement for extreme evaluations in a specified accumulation of forecasts to occur only in such a case where every forecast is formally acknowledged to be fully justified or not justified (the last expression means that the alternate forecast was justified). Cases of partial justification of forecasts may be reduced to a scheme of alternate forecasts.

- 1 -

RESTRICTED

[Main Text]

1. As is generally known, correct forecasts, i.e. scientific predictions, have great agricultural value. Therefore it is of interest to find a correct, scientifically-founded method for evaluating the accuracy of various forecasts supplied. The sought-for evaluation should be an impartial, objective measure of the average result of precise specification of our knowledge concerning the occurrence or non-occurrence of certain facts, which actually take place, if the consumer of forecasts is really guided by the forecasts in his practical activity. Let us note immediately two different parts of the agricultural problem.

The first part consists in improving the knowledge of the consumer on occurrence or non-occurrence of corresponding phenomena. This part depends on the merit of the employed method of forecasting, on the ability of the forecaster to apply this method and on the correct explanation to the consumer of the real meaning of the specified forecast. Unfortunately the importance of the explanation to the consumer is often not given due attention, and the consumer understands the forecast differently from the forecaster and not quite as it should be understood from the point of view of probabilities.

The second part consists in the ability to exploit efficiently the obtained data in agricultural activities and depends on the consumer of forecasts, his technical equipment and the type of his practical activity.

RESTRICTED

RESTRICTED

We consider here only the first part of the problem. In this respect the desired evaluations are supposed to serve as objective measures of an average virtual improvement of the knowledge of consumers on some expected or unexpected occurrences. Such evaluations allow one to correctly compare the merits of various forecasting methods and hence to select correctly that method which indeed provides the consumer with most accurate knowledge, i.e. it represents the best among the methods compared. These evaluations enable one to compare the results of prognostic activities of various persons and institutions. We may also get an answer to the question: in what part of prognostic activity did the scientific development of methods of forecasting attain the highest theoretical level; for instance, whether the theoretical efficiency level will be higher for hydrological prognosis of maximum levels of spring floods or for long range prognoses of precipitations according to the specified physico-geographical region.

2. Let A_1 be prognosis of number 1 and let P_1 be its correct evaluation. The success of a separate individual prognosis is an accidental value, and therefore it is not characteristic of the specified method or prognosis. This explains why the theory leaves it to the compiler of official instructions on evaluation of the successful outcome of a prognosis a certain amount of freedom in the conclusion of formal successes of separate prognoses. The theory does not a priori require the instruction compiler to limit the selection of formal evaluations within 100% range of success, but does demand that this selection should correspond to practical needs and should be devoid of absurdities. A case may be called

- 3 -

RESTRICTED

RESTRICTED

absurd, if a great error in prognosis is made to correspond to a favorable evaluation. However we shall distinguish formal evaluations, designated by the letter a , and considered as initial data, from the correct (desired) evaluations $\bar{r}$.

Of decisive importance for a sufficiently long series of prognoses is the arithmetical average of correct evaluations $\bar{r}_i$ ($i = 1, 2, \dots$). We shall designate this arithmetic average by means of a dash $\bar{r}$. Thereupon the problem is reduced to the question: how should one pass, on a scientific basis, from evaluation a_i , hereafter let by some instruction, to the evaluation $\bar{r}$ should correctly characterize the efficiency of a sufficiently long series of prognoses. In this scheme the evaluations $\bar{r}_i$ have the significance of intermediate results for computation according to initial data of the final result $\bar{r}$. Therefore the principle of the subject will not be altered, if instead of the real evaluation $\bar{r}_i$ we shall first look for the approximate (tentative) values r_i , determined up to the accuracy of certain factor k , the value of which will be known after the volume of accumulation of prognoses, for which the final evaluation of probability $\bar{r}$ is sought, will be accurately defined. In this case $\bar{r} = k\bar{r}$, where k depends on the evaluated accumulation.

2. The initial evaluations a_i and the final evaluation $\bar{r}$ may be expressed on two basic scales. 1) On the usual percent scale, where a 100% evaluation corresponds to maximum possibility and evaluation 0% corresponds to maximum improbability, i.e. to the highest negative possibility (maximum possibility

- 4 -

RESTRICTED

RESTRICTED

of alternate prognosis). In this case it is advisable to agree that the point of indifference (equal to the concept of improbability) should always coincide with an evaluation of 50%, what may be attained by a corresponding definition of the middle part of the scale. 2) In a standardized scale of evaluations, in which 100% of the first scale corresponds to the evaluation +1, and the evaluation 0 corresponds to the evaluation -1; the point of indifference corresponds to the evaluation 0.

We shall mostly use the standardized scale of evaluations without specially mentioning it. The same values on the percent scale will be supplied with an asterisk for distinction. Therefore the transition from $\bar{a}$ to a or the transition from $\bar{R}$ to R^* will be defined by the formulas

$$a = \frac{\bar{a}^*}{50} - 1; \quad R^* = 50(\bar{R} + 1)\%; \quad -1 \leq a; \quad \bar{R} \leq 1; \quad (1)$$

4. The concept of improbability of the method of forecasting corresponds to such a final evaluation $\bar{R} = 0$ of a sufficiently long series of prognoses, established according to such method, that would be correct in the case of full independence of actual occurrences from forecasted ones. Such independence means that the prognosis does not change the probabilities of corresponding occurrences. If, for instance, the conditional probability of precipitations remains the same in the case of forecast precipitations, as well as in prognosis "without precipitations" then the specified method of precipitation forecast is unsuccessful and its final evaluation $\bar{R}$ for a sufficiently long test series should tend to zero. In such a way we are led to the necessity to express the evaluation by the probability theory.

- 5 -

RESTRICTED

RESTRICTED

Formal evaluations $\bar{a}_1$, if we do not introduce corrections, lead to average formal evaluations $\bar{a}$. The latter do not give the correct characteristic of merits of the forecasting method, as may be seen from the following example. Destructive earthquakes are rather rare phenomena; and although the location of seismically active centers is well known, still the population does not know at all which day such an earthquake will occur. The probability of such an occurrence may be computed with little accuracy; therefore the problem of a successful forecasting seems to be "difficult". On the contrary the prognosis "tomorrow will be clear weather" for Tashkent in July appears to be an "easy" forecast, because the unconditional probability of a clear day is here very high and the population knows it without prognoses. A prognosis alternate to "easy" should of course be "difficult".

In the first case, if among every 100 prognoses of destructive earthquakes only 20% on the average will materialize, then even this evaluation, although formally low, really will mean a definite achievement in the processing of methods of earthquakes forecastings, because such prognoses essentially improve the knowledge of the population concerning the occurrence of forecasted events. In the case of "easy" prognoses, even a high average evaluation of 80% formal correctness may hide a negative characteristic of the forecasting method, if the unsuccessful forecasting method corresponds to a still higher probability of the forecast event.

RESTRICTED

RESTRICTED

It follows from the definition of unsuccess itself that the prognosis may be efficient (with a corresponding sign + or -) only if it changes probabilities, i.e. when the conditional probabilities are different from unconditional. For this reason we do not refer to information on constant probabilities in various handbooks as a proper prognostic activity, but we consider it a category of general knowledge, pertaining to the field of statistical research. For instance, handbooks on climatology should describe climatic regions, according to months of the year, the probabilities of temperatures, precipitations etc.

In this way the consumer's level of knowledge of prognoses, from the point of view of prognostic activity of the forecasting service, may be established by the knowledge of such probabilities, which may be supplied in the form of regular handbooks, giving details on corresponding physico-geographical regions and presenting tables of variation of probabilities of events according to basic periods (annual rate; diurnal rate; for instance in due cases for forecasts of boreal aurorae, or radio fadeouts variation of probabilities according to phases of 11-year cycle of solar activity etc). Besides we may admit that every one is able by himself, without the help of a forecaster, to establish a short range inertial weather prognosis: "tomorrow we expect the same weather as today" and to explicit the amount of truth included in it. The probability of agreement of such prognoses with their annual rate could have been computed and published. The fact that such handbooks are not compiled has not been of basic importance, because this is a temporary phenomenon, a deficiency of the present day, quite expendable if desired.

- 7 -

RESTRICTED

RESTRICTED

5. The activity result of the forecasting service may and should be defined by current corrections of data of this kind in fixed handbooks, supplemented by the forecasting service. Let A_1 be some text of forecasting and C an event mentioned in it, which has an unconditional probability p according to the handbook for the specified object and time, but in the presence of the prognosis A_1 its conditional probability is different; let us designate it by p' . It is obvious that the consumer will fully and correctly understand the sense and value of the prognosis A_1 if for each event C of interest to him, he will be informed on probabilities p and p' (before and after the forecast). If C means "precipitations", and the text of the forecast A_1 states: "clear, without precipitations", then the text corresponds to the prognosis result, if $p' < p$; in the opposite case it does not correspond, and it is more correct to understand the prognosis in the opposite sense. Further without knowing how far p' differs from p , the farmer lacks a reliable source to judge how far his preventive measures will warrant his agricultural expenses.

6. As we see, the evaluation of forecasts should be based on corresponding probabilities. They should, as it is said, be stochastically founded. (Note: The word stochastical is of Greek origin and refers to estimates. Therefore "stochastically-based" means theoretically based on analysis, or study, of probabilities). Correct evaluations should take into account, as mentioned before, "difficulty" or "easiness" of various forecasts, and these concepts are also defined by probabilities. Whatever the material in

- 8 -

RESTRICTED

RESTRICTED

the forecasting text may be (on river levels or on approach of frosts, etc) the theory of evaluation deals only with corresponding probabilities. On the other hand, for a forecast of the same kind, as for example "rain with change into snow" the probabilities p vary during the year, and therefore the "difficulty" of the forecast varies too. Therefore if for the determination of the final evaluation $\bar{F}$ we will be able to overcome the difference in "difficulties" of forecasts, at least within the frame of a permanent text of forecast (for this purpose the evaluations a_1 of forecasts of various difficulties should be reduced to evaluations r_1 of forecasts of certain "standard" or "uniform" difficulty), this will solve the problem of building a universal scale of comparison of evaluations of arbitrary forecasts.

7. Let us have a forecast with a certain text A_γ , which predicts the accomplishment of an event C , with a unconditional probability p . Then the alternate forecast which we shall designate with a dash above $\bar{A}_\gamma$, stating that the event C will not occur, has a "difficulty" determined by the probability $1 - p = q$. If the event C fully occurred, then the formal probability a_1 of the first forecast equals $+1$, and for the second forecast we have $a_2 = -1$, and the contrary respectively if C did not occur. We have to pass from the formal evaluation a of forecast A_γ to the tentatively evaluation r_γ , corrected by the "difficulty" of the forecast. The magnitude r is the desired function of a and p , or what is the same, of a_γ and T'_γ , where T'_γ is the measure of "difficulty", which has

- 9 -

RESTRICTED

RESTRICTED

to be defined by the probability p . Therefore the choice of the form of the expression of the difficulty T' by p is not essential, but depends on our present. For considerations of symmetry let us first admit for the difficulty T' and the easiness L' the following designations: $T' = a/p$; $L' = p/a$. In this way the difficulty T' and the easiness L' are connected by the relation $T'L' = 1$. The difficulty T' of the forecast A is the easiness L' of the alternate forecast $\bar{A}$, and vice-versa. It is also obvious that in the case of a correct evaluation of the value r , to the positive evaluation $r > 0$ for the realized forecast A , an equal and opposite, in sign, evaluation $(-r)$ should correspond with the respectively not realized alternate forecast $\bar{A}$ and vice-versa.

Below we shall make use of some different definition of difficulty T and easiness L , designating them for distinction without apostrophe. They will be related to T' and L' by

$$T = 1/2 (T' + 1); \quad L = 1/2 (L' + 1); \text{ hence} \quad (2)$$

$$T = \frac{1}{2p}; \quad L = \frac{1}{2q}; \quad L = \frac{T}{2T-1}; \quad T = \frac{L}{2L-1};$$

and as far as the probabilities p and q are positive and smaller than unity, the new magnitudes T and L are always > 0.5 .

8. In the theory of probability often analogy with gambling is made. Let therefore V designate the "gain" of the forecaster if the event C does materialize, and let P be the "gain" of the second gambler if the alternate forecast took place (the event C does not materialize). In the case of

RESTRICTED

RESTRICTED

equality of the forecasting scale and at a constant "bet" to fill the "bank" equal to unity (equal for both chances) we should have $pV = qP = 1$, which corresponds to the classical concept of "harmlessness of play" to the bank. In our scheme the gain V with the sign + (plus) and respectively the "loss" P , with the sign - (minus), means the tentative evaluations (r) corresponding to the formal evaluations of success: $q^* = 100\%$ and $q^* = 0\%$ for the first forecaster.

For a uniform series of similar difficulties in forecasting, the average result $\bar{r} = k\bar{r}$ according to the universal scale will have the evaluation $\bar{r} = k \frac{nB - mP}{n + m}$, where n is the number of justified forecasts and m is the number of unmaterialized events among the total amount $N = n + m$ of uniform forecasts. The concept of unsuccess of the evaluated method of forecasts should correspond to, for $N \rightarrow \infty$, the zero evaluation $\bar{r} = 0$. Since $\frac{n}{N} \rightarrow p$, $\frac{m}{N} \rightarrow q$, and we obtain a second equation: $pV = qP$; here we have for r in the case of a materialized forecast the value: $r = \frac{1}{2p} = 1$, equal to the second definition of difficulty of forecast, and in the case of a unmaterialized forecast, $r = -P = -\frac{1}{2q} = -1$; i.e. it is equal to the easiness of the forecast with the minus sign.

By means of these considerations we fixed the zero of the universal scale of evaluations, making it correspond to the concept of unsuccessful forecast. But we did not yet define the unit of the scale. This will be done at the end by means of the definition of the correcting factor k . For this reason we call the evaluations r tentative.

- 11 -

RESTRICTED

RESTRICTED

9. Let us consider now a more complex case where the instruction provides for more than two stages for the evaluation of materialization of forecasts. Let A_v ($v = 1, 2, \dots, N$) be different and provided by instruction texts of forecasts. To each text A_v corresponds a certain range of materialization of forecast S which is subdivided by the instruction into several non-intersecting sectors $s_{v1}, s_{v2}, \dots, s_{vk_v}$; $k \geq 2$. We shall call these phases of materialization or correctness of forecast. To each phase s_{v1} the instruction allows some formal success of the forecast a_{v1} . In this way for a forecast A_v the materialization of phase s_{v1} means a formal success of forecast s_{v1} according to the specified instruction. Among the number k_v of all phases provided by the instruction for the text A_v of forecast, normally one phase has a maximum evaluation $a_{v1} = +1$ (corresponds to 100% of success); one has a minimum evaluation $a_{vk_v} = -1$ (corresponds to 0% of success); with $k_v \geq 2$, the remaining phases correspond to certain intermediate evaluations of formal success, provided by the instruction.

We may consider it an exceptional case if the compiler of instructions in a certain edition of forecast for some reasons does not establish phases with maximum and minimum evaluation ± 1 ; but this means that he deliberately intends that the final evaluation $\bar{P}$ of forecasting methods should not attain a maximum or minimum evaluation ± 1 . We consider absolutely non-admissible evaluations, surpassing the limits of the interval $(-1, +1)$, for instance "success" exceeding 100%, because this does not correspond to the nature of the problem, according to which 100% should designate the

- 12 -

RESTRICTED

RESTRICTED

maximum probability of success, and the other evaluations the part of the maximum evaluation. (Note: Absurdities of this kind sometimes occur in meteorological literature).

Let the unconditional probabilities of materialization of phases s_{vi} for a specified prognostic region and a month of the year be given by the quantities p_{vi} . Then $q_{vi} = 1 - p_{vi}$ means the probability that the empirical point will not fall into the phase s .

In a general case of multiphase the section s_{vi} may be imagined as the sum of two hypothetical non-intersecting sections: $s_{vi} = s'_{vi} + s''_{vi}$, in which phase s' corresponds to formal success 100%, and phase s'' corresponds to formal success 0%. Let us designate the corresponding probabilities by p'_{vi} and p''_{vi} . We have in a standardized evaluation scale $+1 \cdot p'_{vi} - 1 \cdot p''_{vi} = a_{vi} p_{vi}$. Besides we have $p'_{vi} + p''_{vi} = p_{vi}$. Adding all uniform sections s'_{vi} according to the index i , we obtain a certain region of materialization S'_v , which corresponds to a formal evaluation $a'_v = +1$. The addition of sectors s''_{vi} leads us to the region S''_v with a formal evaluation $a''_v = -1$. We have $S'_v + S''_v = S_v$. These two sub-regions correspond to the probabilities $p'_v = \sum_i p'_{vi}$ and $p''_v = \sum_i p''_{vi}$. It is easy to see that $p'_v + p''_v = 1$ and that $p'_v - p''_v = \sum_i a_{vi} p_{vi}$. Hence it is easy to find p'_v and p''_v . We reduced the multiphase case to the case of two phases S'_v and S''_v and according to the definition of difficulty and easiness of forecasts we have for the forecast A_v :

$$T_v = \frac{1}{2p'_v} = \frac{1}{1 + \sum_i a_{vi} p_{vi}} ; L_v = \frac{1}{2p''_v} = \frac{1}{1 - \sum_i a_{vi} p_{vi}} \quad (3)$$

RESTRICTED

RESTRICTED

* plus part $\delta_2 = 1 - \delta_1$ of forecasts given formal evaluation -1 .

10. In the general case the formal evaluation a_{vi} of the forecast A_v at $-1 \leq a_{vi} \leq +1$ may be considered as a case ~~of~~ part δ_1 of forecasts ~~have been~~ given a formal evaluation -1 . * The average of their evaluations will equal the specified magnitude a_{vi} at $\delta_1 = 1/2 (1 + a_{vi})$. Hence we obtain for r_v the average scale (according to scales δ_1 and

$$\delta_2) \text{ value} \quad r_v = \delta_1 T_v - \delta_2 L_v = 1/2 (1 + a_{vi}) - 1/2 (1 - a_{vi}) L_v \quad (4)$$

11. We now have all data necessary for obtaining the final evaluation r_v . Practically the scene of forecast writing in the record book of forecasts and its evaluation is thought in the following way:

1) General informations: No, v and date of forecast record, code of forecasting method or its alternative, instruction code on evaluation of forecast probability, name of forecaster.

2) Text A_v of forecast, including more accuracy of predicted events in space and time.

3) Weight of specified forecast, η_v .

4) Difficulty T_v of the specified forecast.

5) Readiness L_v of the specified forecast.

6) Virtual materialization of event with derivation of code of the materialized phase S_{vi} according to the actual official instruction of evaluation of forecast.

7) Formal (according to instruction) evaluation a_{vi} of forecast, corresponding to phase S_{vi} and reduced to the standard scale (see above, No 3, formula (1)).

- 14 -

RESTRICTED

RESTRICTED

8) Tentative evaluation r_v computed according to formula (4). The value r_v may be either positive or negative and is written with the corresponding sign.

We add the following explanations to this scheme.

General remark: the form of the book may serve not only for recording official forecasts, but also for recording conditional forecasts, given for the past in order to test the merits of some checked method or its alternative or the merits of an alternate instruction of evaluation. Here the algorithm of compilation of forecast is supposed to be performed so accurately, as to eliminate the element of subjectivity, produced involuntarily if the forecaster knows in advance which of the events, predicted in reversed time did virtually materialize.

To graph 2. It is assumed that the forecasting text does not contain the words "possible" or "probably", as far as the use of such words are considered to pertain not to the content of the forecast, but in the considered cases to data of graph 3.

If the forecaster intends to say that a certain event is not expected everywhere or during the whole predicted period, it may be well expressed by the words "locally" or "temporarily". The case of inserting of such a word in the content of the forecast is considered as a special text of the forecast, for the evaluation of which the instruction should anticipate the definition of corresponding phases of materialization s_{vi} and the corresponding formal evaluations a_{vi} .

RESTRICTED

RESTRICTED

To graph 3. Usually the weight of each forecast is supposed to be unit; in this case the graph is superfluous. We introduce it only to show that the scheme of evaluations is applicable in the case of unbalanced forecasts. The introduction of the scale may prove useful for certain theoretical investigations.

In practical service of forecastings some deviations from a uniform scale may prove adequate in cases when the prognostical situation of the forecaster is unclear (in doubt) or, on the contrary, very distinctive for the construction of a sure forecast. In such cases the text of the forecast may be supplemented with words, signalling the deviation of the forecasting weight from the standard unit weight. The decoding, for instance, may be the following

Content	Weight of Forecast
"possible"	0.5
"Probably"	0.75
"Surely"	1.5
"with great certainty"	2.0

Such decoding (as in general instructions of evaluations) should be familiar to the consumer of forecasts. If forecasts provided with great certainty correspond to a higher virtual correctness, then the average final evaluation of the forecasting activity will be somewhat higher by using such unbalanced scales, and it will be justified as long as the consumer will also get part of the additional useful information.

RESTRICTED

RESTRICTED

To paragraphs 4 and 5. Difficulties and easiness of forecasts are defined according to formula (3). For this purpose it is necessary to know the unconditional probabilities p of materialization of phases, anticipated in official instructions of forecast evaluations. The corresponding probabilities are not known for all stations. But there are ways of their approximation. For stations with a short series of observations stochastic characteristics may be made more accurate by interpolation of the statistical data of stations with many-year observations, taking into consideration the found laws of variation and their dependence on geographical coordinates, orography etc. Use could be made of analogies, questioning of the population, special investigations (for example, for rivers research of traces of high-flood horizons). We assume that the respective investigations are processed and the probabilities p_{vi} defined in some way.

Knowing these possibilities it is easy to figure out the difficulties and easiness. But if we do not know these probabilities, then we are unable to define scientifically even the correct formally true (stochastic) average correctness $\bar{a}_v$ of forecast, anticipated by instruction and corresponding to the specified method of forecasts. Therefore the alternative is the following: either we expect to have scientific, stochastically justified evaluations of applied methods of forecastings and in this case we should define the corresponding probabilities p_{vi} for cases anticipated by actual instruction of evaluations of forecasts (and then there is no difficulty to find objective evaluation of these methods according to the

RESTRICTED

RESTRICTED

universal scale of evaluations) or we renounce the definition of probabilities P_{yi} , but then we also refuse the possibility to have scientifically founded evaluations of the applied methods of forecastings.

In such cases where the probabilities P_{yi} have a pronounced annual behavior it is necessary to detail them as well as the values T_y and L_y according to the months of the year and to apply those values which correspond to the forecasted period.

If these quantities (for instance in the case of disturbances in the ionosphere e) are substantially connected with the 11-year cycle of solar activity, then it is necessary to determine them according to the phase of solar activity in the 11-year period.

If we intend to introduce in short-range forecast of 24 hours corrections in the evaluation of the security of the inertial forecast ("tomorrow will be the same weather as today"), then all possible forecasting situations should be divided into a certain number (10-20) of classes. Possibilities and probabilities should be detailed according to these classes and values which correspond to the situation of the date of the forecast should be applied.

12. Possessing such records in the recording book, we may make various selections in order to find corresponding final average evaluations $\bar{R}$ according to the universal evaluation scale. For example we may select all forecasts made by a certain forecaster in order to obtain an objective evaluation of probability of the forecasting activity of this forecaster.

- 18 -

RESTRICTED

RESTRICTED

We may also in a suitable way search for objective evaluations of probability:

- a) for a given compiled period of service activity of forecasts;
- b) for forecasts corresponding to the given alternate of instruction concerning the comprehension and evaluation of forecasts;
- c) for forecasts corresponding to the given alternate method of forecasting;

d) for forecasts regarding a definite object (for instance a specified river or a specified prognostic region);

e) for forecasts of a definite form (for instance for forecasts of precipitations or for forecasts of frosts or for forecasts with a fixed text) etc.

Let us assume that a suitable selection has been made and the corresponding forecast outlined according to the same form into a separate report. Then the standardization of reports and the search for the final average evaluation $\bar{E}$, generalized for the whole array of forecasts outlined in the report, will proceed in the following way:

According to graph 3 all weights η_v are added. Let us designate the sum by N_0 :

$$N_0 = \sum_v \eta_v$$

According to graph 4 the weighted average difficulty $\bar{T}$ is defined;

$$\bar{T} = \frac{1}{N_0} \sum_v \eta_v T_v$$

RESTRICTED

RESTRICTED

According to graph 5 the weighted average easiness $\bar{I}$ is defined;

$$\bar{I} = \frac{1}{N_0} \sum_v \eta_v L_v$$

According to graph 6 the weighted average evaluation $\bar{F}$ is defined;

$$\bar{F} = \frac{1}{N_0} \sum_v \eta_v r_v$$

The final evaluation $\bar{F}$ of probability of forecasts marked in the report according to the universal scale is determined by the formula:

$$\begin{aligned} \bar{F} &= \bar{F} / \bar{I} & \text{if } \bar{F} \geq 0 \\ \bar{F} &= \bar{F} / \bar{I} & \text{if } \bar{F} \leq 0 \end{aligned} \quad -1 \leq \bar{F} \leq 1 \quad (5)$$

The evaluation $\bar{F}$ is given in the standardized scale (No 3). In order to reduce it to the percent scale we have to use formula (1):

$$\bar{F}_* = 50(1 + \bar{F})\%$$

The evaluation $\bar{F}$ defined in this way is suitable for directly comparing the average probabilities of arbitrary arrays of forecasts.

The evaluation $\bar{F}$ is an average evaluation for an empirical array. Its reliability (correspondence to true significance) depends on the volume of the evaluated array and on the accuracy of the definition of probabilities P_{vi} . The probable error may be figured out according to general rules of statistical mathematics, which we shall not discuss here.

The evaluation $\bar{F}_*$ has the following obvious properties, defining its scientific foundation and practical usefulness.

RESTRICTED

RESTRICTED

a) It reaches the maximum value of 100 if and only if every forecast was defined according to instructions as formally 100% successful.

b) It reaches the minimum value of 0%, if and only if every forecast was defined according to instructions as formally having 0% probability.

c) The true (stochastic) value $\bar{P}_* = 50\%$ (or in the standardized scale to the value $\bar{R} = 0$) corresponds to the concept of theoretical unsuccess of the forecasting method. To such evaluation tends, with N , the evaluation of "forecasts" obtained by means of some kind of lot cry.

d) The procedure given of figuring out an objective evaluation and for reducing formal evaluations to the universal scale of comparison is so flexible that it may be applied to any actual official instruction and to an array of forecasts having various "difficulties".

Submitted

18 October 1949.

- E N D -

- 21 -

RESTRICTED